基于机器学习的快速射电暴搜寻方法综述

刘艳玲[1,2,3,4]，陈卯蒸[1,2,3,4]，袁建平[1,2,3]

（1.中国科学院新疆天文台，新疆 乌鲁木齐 830011；2.中国科学院大学，北京 100049；3.中国科学院 射电天文重点实验室，江苏 南京 210008；4.新疆微波技术重点实验室，新疆 乌鲁木齐 830011）

摘　要：快速射电暴（Fast Radio Burst，FRB）是目前射电天文领域的主要热点前沿。其相关研究被《自然》（Nature）杂志评选为 2020 年十大科学发现之一。FRB 爆发时间极短且鲜少重复的特点，使其观测捕捉到的概率极低。由人工从海量的天文观测数据中识别 FRB 事件是件耗时费力的工作。机器学习技术的蓬勃发展为实时搜寻与多频段联合跟踪观测 FRB 带来了可能。该文从传统机器学习方法和深度学习方法两个方面，对该研究已有的成果进行了分析与总结，并探讨了基于机器学习的 FRB 搜寻技术目前存在的问题和面临的挑战，分析了其未来发展趋势。

关键词：快速射电暴；机器学习；搜寻方法；深度学习；射电天文

中图分类号：TP274　　**文献标识码**：A

快速射电暴（Fast Radio Burst，FRB）是一种持续时间仅为数毫秒的爆发性、高度色散的脉冲式射电辐射天文现象，瞬时辐射流量可达数十央斯基（Jy）[1-6]。2007 年，Lorimer[1]在分析澳大利亚 Parkes 天文台脉冲星巡天的历史数据时，首次发现了这种天文现象。直到 2013 年，Dan Thornton 等人[2]在新的 Parkes 巡天观测中发现了四个不同色散量的射电暴后，首次将该类现象命名为快速射电暴。随后，FRB 作为一种新发现的天体物理现象被广泛接受[7]。快速射电暴是当前研究最多的天体物理瞬变之一，但它们的起源以及是否存在多种类型的前身和发射机制仍是未解决的问题[5]。国内外已有多个团队开展 FRB 观测与研究。今年 2 月 19 日，中国慧眼卫星团队宣布：确认在银河系内发现的快速射电暴 FRB 200428 来源于磁性 SGR J1935+2154[12]。这也是人类首次证明快速射电暴可以起源于磁性爆发。未来，还需要基于大量的 FRB 观测数据进行研究，以回答更多与其起源和发射机理有关的问题。

除了 Parkes 望远镜，国际上已有多个望远镜参与 FRB 的观测与搜寻工作，并发现了 FRB。如美国 Arecibo 望远镜[13]和 GBT[14]，澳大利亚的 UTMOST[15,16]和 ASKAP[17,18]，加拿大的 CHIME[19,20]，德国的 100 米 Effelsberg 望远镜[21]，意大利的 SRT [22]。作为世界面积最大、灵敏度最高的 FAST 也观测到了 FRB 事件[23,24]。从目前的 FRB 事件观测率（受限于观测时间和视场）推断，FRB 事件的发生率非常高，每天有数千起，这意味着宇宙中产生这些辐射源的物体一定很多[5]。预计未来几年，FRB 探测率将迅速提高，甚至每年可探测数百至数千个 FRB。FRB 的研究进程随着 FRB 探测率的快速增加将进入一个新的纪元。因而，如何有效的从海量观测数据中快速筛选出稀有的、真实的 FRB 事件，成为开展 FRB 科学研究首要解决的一个重要问题。

本文第 1 节介绍传统的 FRB 搜寻方法，第 2 节主要阐述和分析基于机器学习的 FRB 搜寻技术，最后探讨了基于机器学习的 FRB 搜寻技术目前存在的问题以及未来发展趋势。

基金项目：国家自然科学基金（11903071）。

作者简介：刘艳玲，女，博士研究生，研究方向：天文技术与方法，深度学习，信号处理。Email:liuyanling@xao.ac.cn

通讯作者：陈卯蒸，男，博士，正高级工程师，研究方向，射电天文技术与方法，接收机技术等。Email:chen@xao.ac.cn

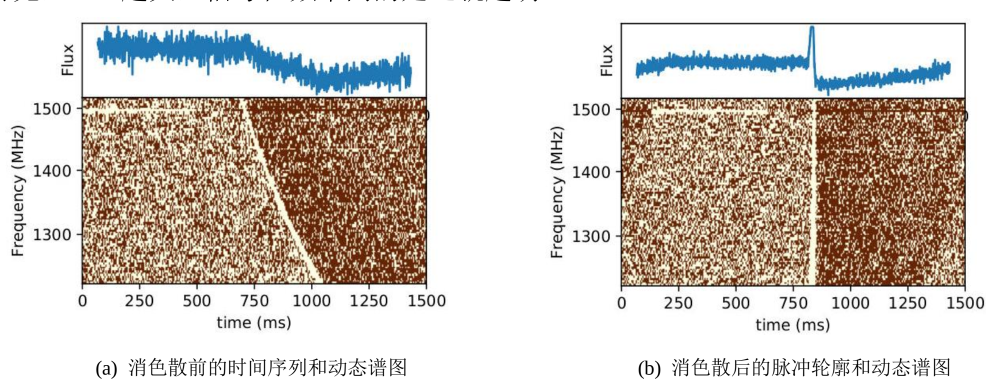
1 传统的快速射电暴搜寻技术

FRB 看起来类似于银河系脉冲星的单脉冲，但是它们的大色散延迟表明他们通常来自银河系外（目前，FRB 200428 除外）。脉冲星辐射极其微弱，但具有非常稳定的周期性，大部分的脉冲星需要通过周期折叠处理才能得到它的积分脉冲轮廓。而 FRB 鲜少周期性重复但明亮的特点，使它与脉冲星的搜寻过程有重复也有不同。

射电信号穿过星际介质到达地球，会发生色散、散射和闪烁现象。色散，散射和闪烁会导致脉冲轮廓的展宽和变形，以及强度的变化。其中色散的影响最大，表现为高频信号比低频信号先到达，如图 1a 所示。因此，消色散是传统脉冲星和 FRB 搜寻过程中的关键技术手段。在 FRB 搜寻过程中，通过试验一系列的色散量（Dispersion Measure，DM）对观测数据进行频率通道数据移动消除色散延迟，再对调整后的频率通道数据叠加生成不同信噪比的时间序列（脉冲轮廓），当最大化信噪比的 DM 出现时，经过消色散处理的动态谱和时间序列成图表现为图 1b。色散延迟 Δt 跟 DM 的关系：

$$\Delta t = 4.15\times(v_2^{-2} - v_1^{-2})\times DM \qquad (1)$$

由此可见，DM 越大，信号在频率间的延迟就越明显。



(a) 消色散前的时间序列和动态谱图　　(b) 消色散后的脉冲轮廓和动态谱图

图 1　Lorimer 发现的 FRB（FRB 010724）

Fig.1 The Lorimer burst

针对较强流量的脉冲星，其单脉冲与 FRB 一样，通过最大化信噪比的 DM 值消色散处理后即可以被发现。但是对于大部分辐射极其微弱的脉冲星信号，经消色散后，还需借助傅里叶变换确定其周期，然后进行信号周期叠加（提高信号比），最后输出候选体的脉冲轮廓、时间-相位图等信息进行确认。

这些罕见且鲜少重复的单脉冲事件是由自动化、高性能的基于消色散理论的软件管道进行搜寻，如 HEIMDALL[25]，FDMT[26]，Bonsai[7]，Amber[27]，CDMT[28]，Presto[29]，BEAR[30]等。FAST 多科学目标同时巡天规划采用的是基于 HEIMDALL 建立的 FAST_Miner 管道[24]，在 20 多个 GPU 服务器上进行 FRB 事件搜寻，生成的候选体经过参数初步过滤后，再由人工进一步的审查。李柯伽团队开发了 Piggyback 后端和 BEAR 搜寻软件，安装在昆明 40 米和新疆南山 26 米望远镜进行 FRB 观测搜寻工作[30]。这些管道对经大量的 DM 值消色散处理后得到的时间序列（脉冲轮廓）进行阈值判断，任何峰值信噪比高于阈值的都被报告为候选体。由于射频干扰（Radio Frequency Interference，RFI）、系统增益变化或者其他的因素，这些算法面临着噪声和 RFI 伪装成 FRB 的挑战，即假阳性：预测为正，实际为负。这样的单脉冲检测程序为了不错过 FRB 事件，制造了数以千计的假阳性候选体。最初，其审查工作是由人工进行

的。但是，随着 FRB 观测数据量的增加，特别是多波束、天线阵所产生的数据呈指数级增长，这也将与假阳性候选体数量的增长相对应。就目前全球 FRB 事件检测到的概率而言，人工筛选 FRB 是一项繁冗、低效率、高成本的棘手工作。

2 基于机器学习的 FRB 搜寻方法

在海量的观测数据中寻找罕见的 FRB，就如同大海捞针。机器学习的应用能够提高搜寻 FRB 事件的速率和准确率。本文根据是否进行人工构造特征，将基于机器学习的 FRB 搜寻技术分为基于传统机器学习的方法与基于深度学习的方法。机器学习在搜寻 FRB 中的应用，实际上解决的是 FRB 和 RFI 或者背景噪声的分类问题。一般使用准确率（Accuracy）、召回率(Recall)、精确率(Precison)等来反应算法的性能。其中，准确率反映算法正确分类正负样本的能力；召回率反应算法正确识别 FRB 的能力，值越高，代表越少错过 FRB 事件；精确率反应被预测为正的样本中 FRB 所占的比例，值越高，代表越少的 RFI 被错分类为 FRB。

2.1 数据集的准备

训练样本是机器学习算法应用的前提。然而，目前得到认证的 FRB 数量非常少，而且可能不是潜在 FRB 群体的代表性样本，因此，不足以为机器学习建立有意义的训练集。见表 1，在传统的机器学习方法中，都是采用来自脉冲星的单脉冲充当 FRB 样本。近几年，在基于深度学习算法的 FRB 搜寻应用中，通常采用模拟 FRB 的方法生成 FRB 样本集，或再补充来自银河系脉冲星的单脉冲充实训练集。因此，目前的机器学习模型，并没有对脉冲星的单脉冲和 FRB 进行区分，而是归为一类，后处理中再根据是否周期性重复以及 DM 值等做进一步的判断。由于 FRB 单脉冲形态简单，用较少的参数即可建立仿真模型。在设计 FRB 仿真算法时，会综合考虑色散、散射和闪烁的影响来模拟 FRB 脉冲信号，然后把它叠加在真实的仅含有背景噪声和干扰的观测数据上生成 FRB 样本。以模拟方式建立 FRB 样本库，可以通过控制参数获得 FRB 脉冲在色散、宽度、幅度及闪烁模式上合理分布的样本库。

表 1 部分样本数据集比较

Table 1 The comparison of some FRB datesets

Type	Author（year）	Telescope	FRB sample source	Dataset size	Positive	Negative
ML	Forster et al[31](2016)	Arecibo	Pulsar	15070	691	14379
	Farah et al [32](2018)	UTMOST	Pulsar	10000	/	/
	Michilli et al [33] (2018)	LOFAR	Pulsar	53066	18003	35063
DL	Connor et al [34] (2018)	WSRT	Sim FRB, Pulsar	21246	Half	Half
	Zhang et al [35] (2018)	GBT	Sim FRB	400000	Half	Half
	Devansh et al [36] (2020)	GBT	Sim FRB, Pulsar, Crab giant pulses	82574	40000	42574

由于 RFI 来源复杂，种类繁多，模拟 RFI 是比较困难的，而且 RFI 时刻存在于观测数据中，因此，在现有的研究中，负样本集均由真实的观测数据产生。这也给机器学习模型识别 RFI 提出了挑战。因为无法控制训练样本集中的 RFI 种类及其数量，会造成某一种或多种类型 RFI 的样本数量偏少，致使模型无法获得拒绝该类 RFI 的能力，从而降低识别率。

需要指出的是无论是来自脉冲星的单脉冲，还是模拟的 FRB 样本，都没有现成可用的公共数据集，大家都是针对不同的望远镜、接收机及观测终端记录的数据进行设计。因此，由于样本数量、样本分布、样本比例、样本质量等因素的不同，算法间不能直接定量比较。

2.2 基于传统机器学习的 FRB 搜寻方法

基于传统机器学习的 FRB 搜寻方法的实现框架如图 2 所示，其需要有经验的专家花大量时间开展特征工程工作，包括特征构建、提取与选择。特征选择在机器学习中占有相当重要的地位。选取较少的，具有明显物理或统计意义的特征，有助于降低计算成本，提高模型开发和训练速度。基于人工提取特征的机器学习方法的优点是计算简单，模型复杂度低，收敛速度快，对硬件要求低。

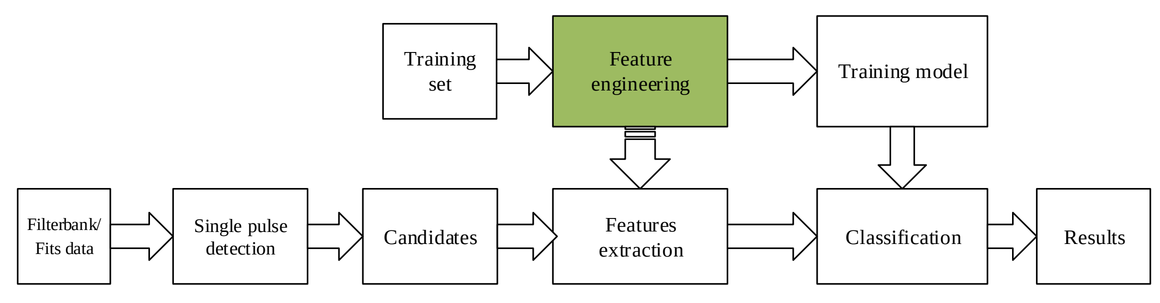


图 2 基于传统机器学习的 FRB 搜寻实现框架

Fig.2 FRB search implementation framework based on traditional machine learning

最早在 2011 年，Thompson 等[37]提出了一种简单的二次判别函数的方法来自动区分噪声、干扰和 FRB。经验证该方法在实践中表现的并不好。因为训练样本是一维时间序列，且信噪比没有设下限。训练好的分类器，在面对新的脉冲强度随时间而漂移的单脉冲时或者强 RFI 时，没有识别能力。

随机森林具有较好抗过拟合能力和处理不平衡数据的能力，在处理大数据集的效率以及对噪声的鲁棒性方面也表现突出，非常适合应用于天文搜寻。因此，随机森林算法在为数不多的 FRB 搜寻应用中受到青睐。Wagstaff 等[38]、Farah 等[32]和 Foster 等[31]均采用了随机森林算法。

Wagstaff 等[38]基于经验，针对候选体的动态谱进行特征设计与提取，二维动态谱相较于一维的时间序列其特征信息丰富且更稳定。选取最小观测频率、DM、信噪比以及候选体事件发生期间、前后区域的图像统计信息等 10 个特征作为模型输入，在 7649 个候选体测试集上，实现了 95.8%的准确率，95.7%的召回率和 97.3%的精确率。

Farah 等[32]在特征使用方面做了一些改进，将候选体输出结果中的信噪比、宽度以及 DM 值作为预分类过滤器的参数对候选体进行第一级的筛选，对于留下的候选体，再从频率-时间数据提取能够表征候选体噪声和信号的 7 类统计特征（如候选体事件窗口以及前后宽度相同窗口的均值和标准差等），输入到机器学习分类器中识别，实现了 98.8%的准确率。这样的两级分类策略，降低了后一级机器学习模型的复杂度，但是对于人类经验依赖程度很高，受限于研究人员的认知水平和经验模式。为了验证系统，他们还对含有 2000 个模拟 FRB 样本的数据集测试，实现了 90%的召回率。值得一提的是，Farah 等[32]开发的是一个低延迟（<24s）的候选体分类管道，实现了近实时分类和电压数据捕获。利用该分类器发现了 FRB 170827，并成功捕捉到了其电压数据，揭示了 FRB 170827 的时间结构。

Foster 等[31]与其他文献中的 FRB 和 RFI 简单的二分类应用不同，其通过人工标记，对 RFI 进一步详

细划分为 8 类，二分类系统在区分单脉冲和这 8 类干扰时表现的很糟糕。因此，针对 RFI 种类多样的特点，建立了多类别概率分类系统。对每个候选体提取了 409 个特征，经模型预处理后减少到 398 个。经测试，该分类模型对单脉冲实现了 96.3%的召回率和 92.35%的精确率。概率多标签分类器的优势是可以根据预测结果，优化人工审查的顺序和时间分配：属于单类别高概率的快速检查，对属于多个类别的进行详细的检查与进一步的分析，这也在一定程度上避免了错失罕见的 FRB 事件。但是，该分类器输入的特征数量比较大，增加了模型的复杂度，也牺牲了速度。

Michilli 等[33]围绕特征选择，开展了细致深入的信息增益价值评估实验，最终选取了脉冲宽，脉冲 DM 的加权平均值，DM-宽度曲线的超额峰度，DM-S/N 曲线的超额峰度以及信噪比这 5 个具有较好区分度的特征。考虑到实际观测数据流中单脉冲罕见的特点，基于 Gaussian Hellinger 快速决策树算法开发了单脉冲分类器，该算法专门用于处理不平衡数据流。与以上应用相比，Michilli 等[33]以最少的输入特征，获得了更好的分类器性能，实现了 98.8%准确率，98.6%的召回率和 98%的精确率。该分类器的过程版本已经在 LOTAAS 观测数据中发现了 7 个新的脉冲星。

2.3 基于深度学习的 FRB 搜寻方法

卷积神经网络的出现，使得深度学习在图像识别领域取得巨大进展。深度学习在 FRB 搜寻中的应用，近几年才开始发展。主要得益于 Farah 等[32]模拟 FRB 样本的启发，解决了深度学习面临的大数据量样本的需求。目前，主要有文献[34-36]的工作

Connor 和 van Leeuwen[34]开发了一个子网络层数较少的多输入深度神经网络架构，将候选体的消色散动态谱、DM-时间阵列、时间序列以及多波束探测信噪比信息作为 2 维 CNN、1 维 CNN 和 FNN 的输入，分别单独提取特征，然后在全连接层整合，最后输出预测结果。该文首次选用模拟的 FRB 充实训练样本，经对 Apertif 的数据测试，该分类器对脉冲星的单脉冲实现了 99.7%的召回率。模拟 FRB 为模型提供了在数量上和样本随机多样性上（如，跨越更大的宽度和 DM 范围）更具优势的训练集，这是以往选用脉冲星的单脉冲作为训练样本无法实现的。另外，该网络的前三个输入之间是高度冗余的，虽然凭经验，三者组合会比单独使用效果更好，但也无疑增加了整个网络架构的复杂性。

Devansh 等[36]基于 8 种深度更深的网络模型（如 VGG16，VGG19，Densenet121/169/201，Xception 等），通过多种组合构建了 11 个二输入的二分类网络架构。其简化了模型的输入，仅将候选体的消色散动态谱和 DM-时间阵列作为输入。经测试，这 11 个分类模型的准确率和召回率都达到了 99.5%以上。该文在 Connor 和 van Leeuwen[34]基础上，做了许多改进：如引入迁移技术训练网络模型，减少了训练参数；采用乘性融合方法组合二输入模型提高网络性能；利用翻转技术增加了 RFI 的样本数量；采取微调技术提高模型在特定应用中的性能等。该算法包已经嵌入到美国 GBT 实时 FRB 搜寻管道中，探测到了 20 颗脉冲星的 2000 多个单脉冲[39]。值得说明的是，虽然网络层数的增加可以提高模型的性能，但是也提高了模型的复杂度和训练难度，另外，层数的增加对模型性能的贡献也可能会达到饱和，甚至会引起梯度不稳定、网络退化、性能下降等问题。如果能对冗余层进行修剪，将有助于提高模型的整体性能。

Zhang 等[35]建立了一个由 17 个卷积层组成的 Resnet 网络架构，以色散动态谱图作为输入，训练数据是由模拟的 FRB 样本和仅含有 RFI 和噪声的观测数据组成。经测试，实现了 88%的召回率和 98%的

精确率。与文献[34]和[36]不同的是，它不依赖于传统消色散技术的搜寻手段，而是将训练好的模型直接用于搜寻原始色散动态谱数据中的 FRB 121102，并成功从 2017 年 8 月 26 日美国 GBT C 波段接收机观测的 FRB 121102 数据中，发现了 93 个 FRB 121102 脉冲。此前，该数据由基于消色散理论的搜寻管道仅检测到 21 次爆发。因此，该方法比传统的消色散搜寻算法在更高的灵敏度、更低的误报率和更快的计算速度方面表现更出色。

基于深度学习的 FRB 搜寻技术，整体性能上较之前有了明显的提升。现有的 3 篇文献均是直接将特征提取的任务交由卷积神经网络来自动完成，这对于深度学习模型计算的复杂度以及训练难度都提出了挑战。在数据预处理过程中，仅对输入网络模型的候选体动态谱图、DM-时间阵列图等做基本的尺寸调整和标准化处理。这在很大程度上减少了数据处理的工作量，同时避免了人工设计、提取特征的不完备性和偏见性。但是，完全依赖网络模型自动提取特征，缺乏可解释性，在训练数据不平衡的情况下会出现种类歧视，如 RFI 的类不平衡问题。另外，深度学习算法对于硬件要求较高，收敛速度缓慢，需要花费较长的时间训练。但是其基于 GPU 的前向传播的推理过程是非常快的，可以满足 FRB 事件在线搜寻的应用需求。

3 问题与展望

基于机器学习的 FRB 搜寻技术发展时间相对较短，相关的研究论文较少，且还没有形成大规模的广泛应用。目前探测到的 FRB 事件大部分是依赖于传统消色散搜寻技术。基于机器学习的 FRB 搜寻技术在研究和应用方面都还有继续改进的空间。主要表现在：

（1）由于望远镜类型（单碟或者阵列），接收机的类型（单波束或者多波束），观测终端以及数据格式（Filterbank，FITS 或者 VDIF）等不同，算法需要做针对性的设计。望远镜所在地的 RFI 环境也会对数据质量产生非常大的影响。因此，提高算法的通用性和泛化能力具有重要的意义。

（2）除 zhang 等[35]外，目前的算法都是针对基于消色散搜寻管道生成的候选体分类。在宽带观测中，脉冲的频率结构是高度可变的[40]，基于消色散数据频率积分时间序列的信噪比区分 FRB 和 RFI，会导致错过 FRB。因此，直接将分类器对原始观测数据流进行 FRB 搜寻，无论是从搜寻速度还是召回率的角度考虑都值得进一步的研究和探索。

（3）目前模型训练所用的样本，几乎都是来自脉冲星的单脉冲或者模拟样本，这会导致训练的模型过拟合脉冲星或者模拟 FRB 的特性。因此，随着探测到的 FRB 样本的增多，以及对其认知和理解的加深，持续优化模拟样本或增加真实观测样本以提高训练样本的质量，对于提高算法的性能具有重要意义。

（4）RFI 环境会持续恶劣，将呈现出更复杂的多样性。甚至一些 RFI 会具有 Perytons 的特性[30,41]。因此，除了采取必要的 RFI 消除和缓解措施以外，通过解决训练样本中 RFI 的类不平衡问题，可以进一步提高算法的鲁棒性。

（5）目前的算法主要以二分类（天体物理现象和非天体物理现象）为主。不同 FRB 之间，FRB 和脉冲星单脉冲之间，脉冲星的单脉冲之间，以及不同类型的 RFI 之间也会存在很多差异。因此，对样本

进行更细致的分类，将有助于进一步提高分类算法的性能。

4 结束语

FRB 搜寻速度与准确率对于实现触发多频段跟踪观测和电压数据转储非常关键。未来是大数据的天下，就目前现有观测条件下的数据量也已远远超过了人力所及。因此，应用机器学习将有助于解决目前人工筛选 FRB 候选体难以为继的现状。在分类器设计方面，鉴于深度学习算法对大数据集卓越的学习能力，以及其避免了人工设计、提取特征的弊端，深度学习在 FRB 搜寻应用方面将会发挥更大作用。

参考文献：

[1] LORIMER D R, BAILES M, MCLAUGHLIN MA, et al. A bright millisecond radio burst of extragalactic origin[J]. Science, 2007, Nov 2;318(5851): 777-780.

[2] THORNTON D, STAPPERS B, BAILES M, et al. A Population of Fast Radio Bursts at Cosmological Distances[J]. Science, 2013, 341(6141): 53-56.

[3] PETROFF E, BARR E D, JAMESON A, et al. FRBCAT: the fast radio burst catalogue. Publications of the Astronomical Society of Australia, 2013, 33: e045.

[4] CORDES J M, CHATTERJEE S. Fast Radio Bursts: An Extragalactic Enigma[J]. Annual Review of Astronomy and Astrophysics, 2019, 57: 417-465.

[5] PETROFF E; HESSELS J W T, LORIMER D R. Fast radio bursts[J]. Astronomy and Astrophysics Review, 2019, 27: 4.

[6] AMIRI M, ERSEN B C, BANDURA K M, et al. Periodic activity from a fast radio burst source[J]. Nature, 2020, 582(7812): 351-355. doi: 10.1038/s41586-020-2398-2.

[7] AMIRI M, BANDURA K, BERGER P, et al. The CHIME Fast Radio Burst Project: System Overview[J]. The Astrophysical Journal, 2018, 863(1):48-75.

[8] DECOENE V, KOTERA K, SILK J. Fast radio burst repeaters produced via Kozai-Lidov feeding of neutron stars in binary systems[J]. Astronomy & Astrophysics, 2021, 645: A122.

[9] KATZ J I. Fast radio bursts[J]. Progress in Particle and Nuclear Physics, 2018, 103: 1–18.

[10] ZHANG, B. The physical mechanisms of fast radio bursts[J]. Nature, 2020, 587: 45–53.

[11] XIAO DI, WANG FAYIN, DAI ZIGAO. The physics of fast radio bursts[J]. Science China-Physics Mechanics & Astronomy, 2021, 64(4): 249501.

[12] LI C K, LIN L, XIONG S L, et al. HXMT identification of a non-thermal X-ray burst from SGR J1935+2154 and with FRB 200428[J]. Nature Astronomy, 2021, 5(4): 378–384.

[13] SPITLER L G, CORDES J M, HESSELS J W T, et al. Fast Radio Burst Discovered in the Arecibo Pulsar ALFA Survey[J]. Astrophysical Journal, 2014, 790(2): 85-154.

[14] MASUI K, LIN H H, SIEVERS J, et al. Dense magnetized plasma associated with a fast radio burst[J]. Nature, 2015, 528(7583): 523–525.

[15] CALEB M, FLYNN C, BAILES M, et al. Fast Radio Transient searches with UTMOST at 843 MHz[J]. MNRAS, 2016, 458(1): 718–725.

[16] FARAH W, FLYNN C, BAILES M, et al. Five new real-time detections of Fast Radio Bursts with UTMOST[J]. MNRAS, 2019, 488(3): 2989-3002.

[17] BANNISTER K W, SHANNON R M, MACQUART J P, et al. The detection of an extremely bright fast radio burst in a phased array feed survey[J]. Astrophysical Journal Letters, 2017, 841(1): L12.

[18] SHANNON R M, MACQUART J P, BANNISTER K W. et al. The dispersion–brightness relation for fast radio bursts from a wide-field survey[J]. Nature ,2018, 562(7727), 386–390.

[19] BOYLE P C, and CHIME/FRB COLLABORATION. First detection of fast radio bursts between 400 and 800 MHz by CHIME/FRB[R]. The Astronomer's Telegram 11901,2018.

[20] AMIRI M, BANDURA K, BHARDWAJ M, et al. Observations of Fast Radio Bursts at Frequencies Down to 400 MHz[J]. Nature, 2019, 566(7743):230-251.

[21] MARCOTE B, NIMMO K, HESSELS J, et al. A repeating fast radio burst source localized to a nearby spiral galaxy[J]. Nature, 2020, 577(7789): 190-194.

[22] PILIA M, BURGAY M, POSSENTI A, et al. The lowest frequency Fast Radio Bursts: Sardinia Radio Telescope detection of the periodic FRB

180916 at 328 MHz[J]. Astrophysical Journal Letters, 2020, 896(2): L40.
[23] ZHU W W, LI D, LUO R, et al. A Fast Radio Burst Discovered in FAST Drift Scan Survey[J]. Astrophysical Journal Letters, 2020, 895(1):L6.
[24] NIU C H , LI D , LUO R , et al. CRAFTS for Fast Radio Bursts Extending the dispersion-fluence relation with new FRBs detected by FAST[J]. Astrophysical Journal Letters, 2021, 909(1):L8.
[25] BARSDELL B R, BAILES M, BARNES D G, et al. Accelerating incoherent dedispersion[J]. MNRAS, 2012, 422(1): 379–392.
[26] ZACKAY B, OFEK E O. An accurate and efficient algorithm for detection of radio bursts with an unknown dispersion measure, for single dish telescopes and interferometers[J]. Astrophysical Journal, 2017, 835(1): 11.
[27] SCLOCCO A, LEEUWEN J V, BAL H E, et al. Real-time dedispersion for fast radio transient surveys, using auto tuning on many-core accelerators[J]. Astronomy and Computing, 2016, 14 : 1-7.
[28] BASSA C G, PLEUNIS Z, HESSELS J. Enabling pulsar and fast transient searches using coherent dedispersion[J]. Astronomy and Computing, 2017, 18 :40-46.
[29] RANSOM S M. New Search Techniques for Binary Pulsars. PhD thesis. Harvard Univ, 2001.
[30] MEN Y P , LUO R , CHEN M Z, et al. Piggyback search for fast radio bursts using Nanshan 26 m and Kunming 40 m radio telescopes – I. Observing and data analysis systems, discovery of a mysterious peryton[J]. MNRAS, 2019, 488(3): 3957-3971.
[31] FOSTER G, KARASTERGIOU A, GOLPAYEGANI G, et al. ALFABURST: A commensal search for fast radio bursts with Arecibo[J]. MNRAS, 2018, 474(3): 3847-3856.
[32] FARAH W, FLYNN C, BAILES M, et al. FRB microstructure revealed by the real-time detection of FRB170827[J]. MNRAS, 2018, 478, 1209–1217.
[33] MICHILLI D, HESSELS J, LYON R J, et al. Single-pulse classifier for the LOFAR Tied-Array All-sky Survey[J]. MNRAS, 2018, 480(3): 3457-3467.
[34] CONNOR L, LEEUWEN J V. Applying Deep Learning to Fast Radio Burst Classification[J]. Astronomical Journal, 2018, 156(6): 256.
[35] ZHANG Y G, GAJJAR V, DFOSTER G, et al. Fast Radio Burst 121102 Pulse Detection and Periodicity: A Machine Learning Approach[J]. Astrophysical Journal, 2018, 866(2): 149.
[36] DEVANSH A, KSHITIJ A, SARAH B S, et al. FETCH: A deep-learning based classifier for fast transient classification[J]. MNRAS, 2020, 497(2):1661-74.
[37] THOMPSON D R, WAGSTAFF K L, BRISKEN W F, et al. Detection of fast radio transients with multiple stations: a case study using the Very Long Baseline Array[J]. Astrophysical Journal, 2011, 735(2): 395-402.
[38] WAGSTAFF K L, TANG B, THOMPSON D R, et al. A Machine Learning Classifier for Fast Radio Burst Detection at the VLBA[J]. Publications of the Astronomical Society of the Pacific, 2016, 128(966): 084503.
[39] AGARWAL D, LORIMER D R, SURNIS M P, et al. Initial results from a real-time FRB search with the GBT[J]. MNRAS, 2020, 497(1): 352-360.
[40] GAJJAR V, SIEMION A P V, PRICE D C, et al. Highest Frequency Detection of FRB 121102 at 4-8 GHz Using the Breakthrough Listen Digital Backend at the Green Bank Telescope[J]. Astrophysical Journal, 2018, 863(1):2-10.
[41] FOSTER G, KARASTERGIOU A, GEYER M, et al. Verifying and reporting Fast Radio Bursts[J]. MNRAS, 2018, 481(2):2612-2627.

A Review of Fast Radio Burst Search Methods Based on Machine Learning

LIU Yanling[1,2,3,4], CHEN Maozheng[1,2,3,4], YUAN Jianping[1,2,3]

（1.Xinjiang Astronomical observatory, Chinese Academy of Sciences, Urumqi 830011, China; 2.University of Chinese Academy of Sciences, Beijing 100049, China; 3.Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008, China; 4.Xinjiang Key Laboratory of Microwave Technology, Urumqi 830011, China）

Abstract: Fast radio bursts (FRBs) are a hot topic in the field of astronomy at present. Its related research was also selected by the journal Nature as one of the top 10 scientific discoveries of 2020. The characteristic that FRBs are millisecond-duration and rarely repeated make them hard to be captured. Identifying FRBs from massive astronomical observation data by human review is a time-consuming and laborious task. With the rapid development of machine learning technology, it is possible to carry out a real-time search and multi-frequency tracking for FRB events. This paper analyzes and summarizes the existing representative results from two

aspects: traditional machine learning method and deep learning method. Finally, the existing problems and challenges of FRB search technology based on machine learning are discussed, and future development trend is also analyzed.

Key word: Fast radio burst; Machine learning; Search method; Deep learning; Radio astronomy